MNRAS **468,** 2836–2849 (2017)
Advance Access publication 2017 March 14

doi:10.1093/mnras/stx621

Addressing the too big to fail problem with baryon physics and sterile neutrino dark matter

Mark R. Lovell,[1,2,3★] Violeta Gonzalez-Perez,[4] Sownak Bose,[5] Alexey Boyarsky,[2] Shaun Cole,[5] Carlos S. Frenk[5] and Oleg Ruchayskiy[6]

[1]*GRAPPA, Universiteit van Amsterdam, Science Park 904, NL-1098 XH Amsterdam, the Netherlands*
[2]*Instituut-Lorentz for Theoretical Physics, Niels Bohrweg 2, NL-2333 CA Leiden, the Netherlands*
[3]*Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany*
[4]*Institute of Cosmology and Gravitation, University of Portsmouth, Dennis Sciama Building, Portsmouth PO1 3FX, UK*
[5]*Institute for Computational Cosmology, Durham University, South Road, Durham DH1 3LE, UK*
[6]*Discovery Center, Niels Bohr Institute, Blegdamsvej 17, D-2100 Copenhagen, Denmark*

Accepted 2017 March 10. Received 2017 February 13; in original form 2016 November 7

ABSTRACT

N-body dark matter simulations of structure formation in the Λ cold dark matter (ΛCDM) model predict a population of subhaloes within Galactic haloes that have higher central densities than inferred for the Milky Way satellites, a tension known as the 'too big to fail' problem. Proposed solutions include baryonic effects, a smaller mass for the Milky Way halo and warm dark matter (WDM). We test these possibilities using a semi-analytic model of galaxy formation to generate luminosity functions for Milky Way halo-analogue satellite populations, the results of which are then coupled to the Jiang & van den Bosch model of subhalo stripping to predict the subhalo $V_{\rm max}$ functions for the 10 brightest satellites. We find that selecting the brightest satellites (as opposed to the most massive) and modelling the expulsion of gas by supernovae at early times increases the likelihood of generating the observed Milky Way satellite $V_{\rm max}$ function. The preferred halo mass is $6 \times 10^{11}\,{\rm M}_{\odot}$, which has a 14 per cent probability to host a $V_{\rm max}$ function like that of the Milky Way satellites. We conclude that the Milky Way satellite $V_{\rm max}$ function is compatible with a CDM cosmology, as previously found by Sawala et al. using hydrodynamic simulations. Sterile neutrino-WDM models achieve a higher degree of agreement with the observations, with a maximum 50 per cent chance of generating the observed Milky Way satellite $V_{\rm max}$ function. However, more work is required to check that the semi-analytic stripping model is calibrated correctly for each sterile neutrino cosmology.

Key words: Local Group – dark matter.

1 INTRODUCTION

The properties of the satellite galaxies of the Milky Way offer an opportunity to study the process of galaxy formation and the nature of dark matter. They are among the intrinsically faintest galaxies that have been observed, and thus constitute an 'extreme laboratory' in which to examine the interplay between the underlying cosmological model and astrophysical processes. One property that has been of particular interest is the central density of these objects. The likely distribution of densities – or the more observationally accessible central velocity dispersions – can be predicted from simulations of Milky Way analogue systems using a combination of the satellites' density profiles and mass functions.

The ability to compare theoretical predictions with observational measurements was made possible by two, almost simultaneous developments. First, simulations of Milky Way analogue haloes achieved sufficient spatial resolution to resolve the properties of cold dark matter (CDM) subhaloes on scales of ∼100 pc (Diemand, Kuhlen & Madau 2007; Springel et al. 2008), which is smaller than the size of the brightest satellite galaxies. These simulations predicted that the satellites had cuspy density profiles, and that these profiles were better described by the Einasto profile (Navarro et al. 2010) than the $\sim r^{-1}$ profile predicted for isolated haloes (Navarro, Frenk & White 1996b, 1997). Secondly, masses within the half-light radii of the Milky Way satellites were estimated using the methods developed by Walker et al. (2009, 2010) and Wolf et al. (2010) (but see Campbell et al. 2016 for a realistic estimate of the errors). The results of these observations were interpreted by Walker & Peñarrubia (2011) and Gilmore et al. (2007) as evidence that the

★ E-mail: lovell@mpia-hd.mpg.de

satellites had cored, rather than cuspy, profiles and were thus in tension with the CDM simulation results. However, this interpretation remains contentious (Strigari, Frenk & White 2010, 2014).

A second tension between the observations and the theoretical predictions concerns the expected abundances of dense, massive ($>10^{10}\,M_\odot$) dark matter subhaloes around Milky Way hosts. Boylan-Kolchin, Bullock & Kaplinghat (2011, 2012) found that the six Milky Way analogue dark matter simulations of the Aquarius project predicted a population of subhaloes that were too dense and massive to host the brightest observed satellites. This problem was first identified by inferring the central densities of simulated subhaloes from the peak of their circular velocity curves, denoted $V_{\rm max}$ (Boylan-Kolchin et al. 2011), and persisted when the highest resolution simulations were used to measure the central densities directly (Boylan-Kolchin et al. 2012). The masses of these simulated subhaloes were large enough to guarantee that gas should collapse within them and form a comparatively bright satellite galaxy that should have been detected by satellite galaxy searches (Parry et al. 2012). This issue became known as the 'too big to fail' problem.

Proposed solutions to this problem have adopted at least one out of two approaches. The first is to decrease the number of massive satellites around the Milky Way. This has been achieved for satellites with $V_{\rm max} > 30\,{\rm km\,s^{-1}}$ by invoking a relatively low mass for the Milky Way halo (Wang et al. 2012; Cautun et al. 2014); less massive ($V_{\rm max} < 20\,{\rm km\,s^{-1}}$) subhaloes are prevented from forming galaxies by reionization and supernova feedback (e.g. Bullock, Kravtsov & Weinberg 2000; Benson et al. 2002; Sawala et al. 2016a). These models also predict scatter in the luminosity–mass relation of galaxies; thus, the brightest galaxies need not reside in the most massive haloes, as seen in observations (Guo et al. 2015).

The second approach is to appeal to baryon effects. One possibility is that adiabatic contraction of the gas initially draws dark matter to the halo centre, only to be evacuated violently when supernova feedback occurs (Navarro, Eke & Frenk 1996a; Pontzen & Governato 2012), although if the feedback is too weak then adiabatic contraction of the gas can increase the density of simulated galaxies still further and thus make the discrepancy with observations even worse (Di Cintio et al. 2013). Another possibility is that early feedback from reionization and supernovae lowers the baryonic mass of the halo, so that less mass accretes on to the halo at later times and the redshift zero mass is smaller than in the pure *N*-body prediction (Sawala et al. 2013, 2016b). A third possibility is for tides to remove material from satellites (Fattahi et al. 2016). In practice, these methods also reduce the total mass of the satellites and can bias galaxy formation efficiency such that some late-forming massive subhaloes host relatively faint galaxies (Sawala et al. 2016b). A fourth solution, which affects both the abundance and density of the most massive haloes, is a revision of the cosmological parameters. Polisensky & Ricotti (2014) argued that better agreement with observations was achieved with the cosmological parameters from the *Wilkinson Microwave Anisotropy Probe* (*WMAP*) 3-yr results than with the 1-yr values used in the original Aquarius simulations, due to the lower value of the power spectrum normalization, σ_8. The result is that gravitational collapse begins at an epoch when the mean density of the Universe is lower.

An alternative set of solutions invokes alterations to the dark matter model. It has been shown that a velocity-dependent self-interacting dark matter model successfully evacuates the right amount of dark matter from the subhalo centre, even creating a core as suggested by Walker & Peñarrubia (2011), while remaining in agreement with bounds on dark matter self-interactions set by halo shapes (Vogelsberger, Zavala & Loeb 2012; Zavala, Vogelsberger & Walker 2013; Cyr-Racine et al. 2016; Vogelsberger et al. 2016). Another possibility is for the dark matter to interact with radiation, which also dilutes the central dark matter density (Schewtschenko et al. 2016). A third possibility is that the dark matter is a warm dark matter (WDM) particle, such as the resonantly produced sterile neutrino (Dodelson & Widrow 1994; Shi & Fuller 1999; Dolgov & Hansen 2002; Asaka & Shaposhnikov 2005) that may have already been detected in its X-ray decay channel (e.g. Bulbul et al. 2014; Boyarsky et al. 2014). WDM particles free stream out of small perturbations in the early Universe. This phenomenon reduces the abundance of 10^9–$10^{10}\,M_\odot$ haloes and delays the collapse of those that do form, to an epoch when the Universe is more diffuse and thus the haloes are less dense (Lovell et al. 2012). The creation of a core due to primordial velocities does not help because these are predicted to be smaller than $\sim$1 pc and therefore not relevant for the satellite internal kinematics (Macciò et al. 2012, 2013; Shao et al. 2013).

The challenge of analysing all of these possibilities, some of which are in competition and others complementary to one another, is compounded by stochastic effects. Even within models restricted to CDM, which do not include baryonic processes, large statistical uncertainties are introduced by the stochastic formation of Milky Way like haloes and uncertainty in the Milky Way halo mass, which is expected to be in the range $[0.5, 2.0] \times 10^{12}\,M_\odot$ (Kahn & Woltjer 1959; Sales et al. 2007a,b; Li & White 2008; Busha et al. 2011; Deason et al. 2012; Wang et al. 2012; Boylan-Kolchin et al. 2013; González, Kravtsov & Gnedin 2013; Cautun et al. 2014; Peñarrubia et al. 2014; Piffl et al. 2014; Wang et al. 2015; Peñarrubia et al. 2016). In order to take account of these effects, Jiang & van den Bosch (2015) computed $\sim$10 000 merger trees of Milky Way analogue CDM haloes of a range of masses using a Monte Carlo (MC) method. They then used a semi-analytic model of subhalo stripping (Jiang & van den Bosch 2016) to calculate the $V_{\rm max}$ functions of each halo realization. They found the Milky Way system of satellites, as defined by the inferred Milky Way satellite $V_{\rm max}$ function with $V_{\rm max} > 15\,{\rm kms^{-1}}$, to be a $\sim$1 per cent outlier of the MC-generated distributions.

In this paper, we also use a MC approach to investigate the $V_{\rm max}$ function. Our method, however, differs from that of Jiang & van den Bosch (2015) in many respects:

(i) We use the *ab initio* semi-analytic galaxy formation model, GALFORM to populate haloes and subhaloes with galaxies. In this way, we can select satellites that are luminous, and in particular those with the highest luminosities.
(ii) We apply a correction for baryonic effects which changes the satellite $V_{\rm max}$ values derived from hydrodynamical simulations.
(iii) We make use of new $V_{\rm max}$ estimates for the Milky Way satellites based on the results of hydrodynamic numerical simulations (Sawala et al. 2016b).
(iv) We apply the method to a series of WDM models, specifically a range of sterile neutrino models whose decay is a plausible source of the recently discovered 3.5 keV line (e.g. Boyarsky et al. 2014; Bulbul et al. 2014).

This paper is organized as follows. In Section 2, we describe our methods. These include the generation of merger trees, the population of these merger trees with galaxies, the algorithm for comparing these galaxies with observations and a discussion of the sterile neutrino models used. We present our results in Section 3, and draw conclusions in Section 4.

2 METHODS

The goal of this study is to generate populations of satellite galaxies, including their luminosities and $V_{\rm max}$, for a range of dark matter halo masses and dark matter models, and then compare the results to the measured $V_{\rm max}$ of the Milky Way satellites. We first discuss our semi-analytic model of galaxy formation, and then our implementation of the algorithm for calculating the stripping of satellites galaxy haloes. We then present a brief discussion of the observational data, and end with a presentation of the statistic with which we compare the simulated and Milky Way satellite $V_{\rm max}$ distributions. We end in Section 2.5 by expanding our analysis to WDM with a presentation of our sterile neutrino models.

2.1 Semi-analytic model of galaxy formation

In this section, we describe how we generate merger trees for dark matter haloes, and populate the subhaloes with galaxies by means of a semi-analytic model.

In order to produce populations of satellite galaxies, we generate 5000 halo merger trees using the algorithm introduced by Parkinson, Cole & Helly (2008, PCH), itself an evolution of the extended Press–Schechter algorithm (Bond et al. 1991) for combinations of a dark matter model and a central halo mass. We have selected 14 host halo masses in the range $[0.5, 1.8] \times 10^{12}\,{\rm M}_{\odot}$, and for most of this paper we focus on three in particular: 0.5×10^{12}, 1.0×10^{12} and $1.4 \times 10^{12}\,{\rm M}_{\odot}$. This method is modified for the sterile neutrino models to incorporate a sharp k-space filter, as opposed to the standard real space top hat filter, because the latter introduces spurious haloes at small scales (Benson et al. 2013; Schneider, Smith & Reed 2013; Lovell et al. 2016b).

The merger trees are then populated with galaxies by means of the GALFORM semi-analytic model of galaxy formation (Cole et al. 2000). In this study, we use a variation of the model described in Lacey et al. (2016), run on dark matter merger trees produced assuming a Planck cosmology (Planck Collaboration XVI 2014). When changing cosmologies, some of the model free parameters needed to be changed in order to still recover a good agreement with the set observations used during its calibration (as described in Lacey et al. 2016). We refer to this model hereafter as LC16. The features of this model include star formation, supernova feedback and dynamical friction in the mergers of galaxies. A full description of the model run assuming an underlying Planck cosmology will be presented in Baugh et al. (in preparation) Leading semi-analytic models such as this enable us to attach luminosities to the PCH haloes and subhaloes, and thus develop $V_{\rm max}$ functions for sets of satellites for which their observations can be reasonably assumed to be complete.

Semi-analytic galaxy formation models vary in their predictions for the galaxy population, in particular for satellite galaxies. We therefore also employ a second version of GALFORM as published in Guo et al. (2016), (hereafter referred to as G16) to demonstrate the uncertainties arising from the galaxy formation model; a full description of this model will be presented in Baugh et al. (in preparation). This model differs from LC16 in two ways that are of interest to this study: the feedback in small galaxies is weaker, and the initialization of orbits is different. In order to show the effects of these two model features, we also consider a hybrid model in which the satellite orbits are initialized in the same way as in LC16 but all other features and parameters are drawn from G16; we label this model as G16-2. Both LC16 and G16 have also been recalibrated relative to the models published in Lacey et al. (2016) and Guo et al. (2016) to take account of a satellite merging model developed by Campbell et al. (2015) and Simha & Cole (2016). However, this merging model is not used here because it requires N-body merger trees as an input. Details will be presented in Baugh et al. (in preparation) and Gonzalez-Perez et al. (in preparation). A more careful study would ensure that the parameters are recalibrated self-consistently for the merging model and the cosmological parameters: we differ this work to a future study.

Given the choice of LC16 and G16 for our fiducial model, we select LC16 because it fits a wider range of astronomical observables and in particular gives a better fit to the satellite luminosity function (see Fig. A1). We consider the impact of the change in models in Appendix A. For the remainder of this paper, we use the LC16 model except where explicitly stated otherwise. For all of our models we use the Planck cosmological parameters: $h = 0.6777$, $\Omega_0 = 0.304$, $\Omega_\Lambda = 0.696$, $n_{\rm s} = 0.9611$, $\sigma_8 = 0.8288$ (Planck Collaboration XVI 2014).

The application of the both versions of GALFORM has to be adjusted for the purposes of WDM models. We discuss this issue in Section 2.5.

2.2 From $V_{\rm vir}$ at infall to $V_{\rm max}$ at $z = 0$

The PCH algorithm calculates the number of haloes of a given virial mass and virial circular velocity merging on to a host halo at a given redshift, $z_{\rm infall}$. Two properties that are not predicted by the PCH algorithm are the maximum circular velocity of the object (which is distinct from the virial circular velocity) and the dark matter mass-loss of that object. In this section, we discuss the derivation of these quantities.

We begin with the conversion from virial circular velocities, $V_{\rm vir}$, to maximum circular velocities, $V_{\rm max}$. These two quantities are related by the equation:

$$V_{\rm max} = 0.465 V_{\rm vir} \sqrt{\frac{c}{\ln(1+c) - c/(1+c)}}, \quad (1)$$

where c is the Navarro–Frenk–White (NFW; Navarro et al. 1996b, 1997) profile concentration of the halo as calculated from the halo mass–concentration relation by the GALFORM code at the halo formation time.

One needs to take account of the effects of baryons on the halo mass and $V_{\rm max}$. Sawala et al. (2016b) showed that the isolated dwarf haloes experienced a decrease in their $V_{\rm max}$ relative to dark matter-only simulations due to the expulsion of gas, an effect not included in the collisionless PCH formalism. They showed that the average magnitude of this suppression, $p = V_{\rm max,\,SPH}/V_{\rm max,\,DMO}$ takes the following form:

$$p = \begin{cases} 0.87 & 0 \leq V_{\rm max,DMO} < 30\,{\rm km\,s^{-1}} \\ g \log_{10}(V_{\rm max,DMO}) + c & 30 \leq V_{\rm max,DMO} < 120\,{\rm km\,s^{-1}} \\ 1 & 120 \leq V_{\rm max,DMO}, \end{cases} \quad (2)$$

where g and c are the constants required to make the relation continuous; a similar relation has been determined for more massive haloes by Schaller et al. (2015). The $V_{\rm max}$ and virial mass m are thus adjusted to $V_{\rm max} = pV_{\rm max,\,PCH}$ and $m = p^2 m_{\rm PCH}$, where PCH denotes the values output by the PCH algorithm. We present results in which this modification is both present and omitted in order to show the impact of supernova feedback on the fit to the observed $V_{\rm max}$ function. We also assume that the concentration of the halo is unaffected by this alteration, and that the stripping procedure developed by Jiang & van den Bosch (2016) is still an accurate model for the subhalo mass evolution.

In order to calculate the $z = 0$ $V_{\rm max}$ functions for our populations of satellites we implement the method of Jiang & van den Bosch (2016). The rate of mass-loss for the satellite, $\dot{m}$, at a time t after accretion, is assumed to be given by the equation:

$$\dot{m} = A\frac{m(t)}{\tau_{\rm dyn}}\left(\frac{m(t)}{M(z)}\right)^{\alpha}, \quad (3)$$

where A and α[1] are parameters to be fitted from N-body simulations, $\tau_{\rm dyn}$ is the dynamical time, $m(t)$ is the mass of the subhalo at time t. $M(z)$ is the mass of the host halo at redshift z, and is calculated using the code developed by Correa et al. (2015a,b,c). Jiang & van den Bosch (2015) and Jiang & van den Bosch (2016) fit $\alpha = 0.07$, and a mean of A, $\bar{A} = 0.86$. They extract sample values of A from a lognormal distribution using this $\bar{A}$ and a standard deviation of 0.17. however, we recalibrate this parameter for our work (see below).

The dynamical time is calculated based on the estimated overdensity of the halo at each redshift, denoted Δ_c. The relationship between the two is

$$\tau_{\rm dyn} = \frac{1.628/h}{\sqrt{\Omega_0(z+1)^3+\Omega_\Lambda}}\left(\frac{\Delta_c}{178}\right)^{-0.5}, \quad (4)$$

and Δ_c itself is given by

$$\Delta_c = 18\pi^2 + 82(\Omega(z)-1) - 39(\Omega(z)-1)^2, \quad (5)$$

where $\Omega(z)$ is value of the cosmological matter density parameter at redshift z, as shown by Bryan & Norman (1998).

The next step is to translate the change in virial mass into a change in $V_{\rm max}$, which is achieved via the relation fitted to simulations in Peñarrubia, Navarro & McConnachie (2008) and Peñarrubia et al. (2010):

$$V_{\rm max}(z=0) = 1.32V_{\rm max}(z=z_{\rm infall})\frac{x^{0.3}}{(1+x)^{0.4}}, \quad (6)$$

where x is the ratio of the redshift zero mass to the infall mass, i.e. $x = m(z=0)/m(z=z_{\rm infall})$.

As a check of our method, we compare the results of our computation to those of N-body simulations. In Fig. 1, we plot the $V_{\rm max}$ functions for subhaloes in the CDM-Copernicus Complexio (COCO) simulation (Hellwing et al. 2016), a zoomed N-body simulation with a high-resolution region of radius $\sim 17\,h^{-1}$Mpc and simulation particle mass of $1.1 \times 10^5\,{\rm M}_\odot$. Here, we include subhaloes out to the radius of spherical top-hat collapse, $r_{\rm th}$, in order to be consistent with the PCH algorithm outputs. We also plot the median of ∼100–700 (highest mass–lowest mass bin) $V_{\rm max}$ functions in which we retain all subhaloes that had an accretion $V_{\rm max}$ greater than $20\,{\rm km\,s^{-1}}$irrespective of whether they host a satellite galaxy, with the exception of those subhaloes that are located within other subhaloes at redshift zero. For this comparison, we also do not apply the Sawala et al. (2016b) correction since COCO is a dark matter-only simulation. We select COCO haloes in the following mass brackets: [0.4, 0.6] $\times$ 10^{12}, [0.9, 1.1] $\times$ 10^{12} and [1.3, 1.5] $\times$ $10^{12}\,{\rm M}_\odot$, and the masses we use are the mass enclosed within $r_{\rm th}$. The PCH masses are drawn from the same brackets in mass, and are selected to fit the halo mass function of Jenkins et al. (2001). In both the N-body subhalo and semi-analytic galaxy cases, we select only objects that are substructures of the host halo rather than substructures of satellites.

[1] Jiang & van den Bosch (2015) denote this parameter as 'γ'. We instead use α in order to avoid confusion with the GALFORM feedback power-law index.

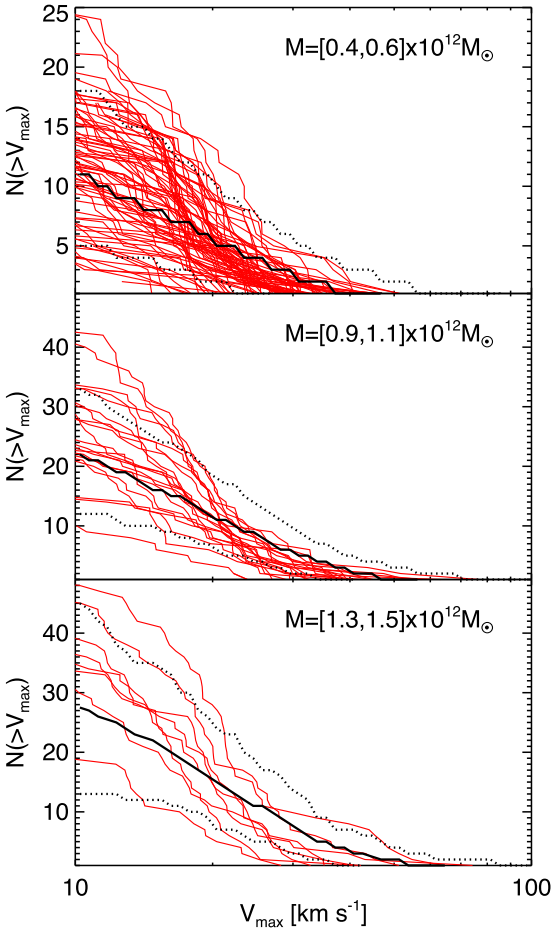

Figure 1. $V_{\rm max}$ functions in CDM N-body simulations and those computed using the PCH + stripping method described in Section 2.2. The COCO $V_{\rm max}$ functions are shown as red lines, and for the PCH functions we show the median $V_{\rm max}$ function across ∼1000 haloes as a solid black line and the 68 per cent scatter regions by dotted lines. In both cases, we restrict the selection to include only subhaloes for which the peak value of $V_{\rm max}$, $V_{\rm peak} > 20\,{\rm km\,s^{-1}}$. The three panels show results for three central halo mass bins: [0.4, 0.6] $\times$ $10^{12}\,{\rm M}_\odot$ (top panel), [0.9, 1.1] $\times$ $10^{12}\,{\rm M}_\odot$ (middle) and [1.3, 1.5] $\times$ $10^{12}\,{\rm M}_\odot$ (bottom). The distribution of PCH halo masses across each bin is determined according to the halo mass function.

There is good agreement between the semi-analytic model and the simulation at $V_{\rm max} > 20\,{\rm km\,s^{-1}}$. In order to achieve this agreement, we recalibrated the $\bar{A}$ parameter to $\bar{A} = 1.4$. The semi-analytic model lacks the satellites with $V_{\rm max} > 80\,{\rm km\,s^{-1}}$, and slightly overpredicts the upward scatter around $V_{\rm max} \sim 24\,{\rm km\,s^{-1}}$ in the middle mass bin. The model consistently underpredicts the number of satellites with $V_{\rm max} < 20\,{\rm km\,s^{-1}}$. Potential causes of this discrepancy include the presence of subsubstructure in the simulation data and a tendency for the model to overstrip small haloes. One should also bear in mind that the COCO simulation was performed with the *WMAP-7* cosmological parameters rather than Planck, therefore a careful study will require that the model be calibrated against simulations using Planck.

2.3 Milky Way satellite properties

Our analysis requires two properties of the Milky Way satellites, their V-band magnitudes and their $V_{\rm max}$ values. We source our V-band values from the data set compiled by McConnachie (2012),

and measured by de Vaucouleurs et al. (1991), Irwin & Hatzidimitriou (1995), Martin, de Jong & Rix (2008). The $V_{\rm max}$ are much more difficult to measure, and typically involve some cross-correlation with CDM simulations. One example of this is the work of Sawala et al. (2016b), who use high-resolution hydrodynamical simulations to derive likely $V_{\rm max}$ values for nine of the dwarf spheroidals based on the simulated satellite's luminosities and central densities. This method has the advantage of selecting the haloes that are most likely to host satellites, whose $V_{\rm max}$ is biased relative to CDM simulation expectations. We therefore use $V_{\rm max}$ plus associated error bars derived from Sawala et al. (2016b) where available. For satellites not included in their study we use the $V_{\rm max}$ values and error bars collated in Jiang & van den Bosch (2015), which were obtained using a likelihood analysis of the satellite velocity-dispersion (Kuhlen 2010; Boylan-Kolchin et al. 2012), and rotation curves (van der Marel & Kallivayalil 2014).

2.4 Likelihood from $V_{\rm max}$ distributions

Here, we summarize the statistical method for comparing the observational and simulated $V_{\rm max}$ distributions. It is identical to that of Jiang & van den Bosch (2015) except where noted below.

The goal is to determine the probability that the Milky Way satellite $V_{\rm max}$ function can be drawn from the distribution of simulated functions. We will establish the statistical scatter between the simulated $V_{\rm max}$ functions, and calculate the mean deviation between the measured Milky Way $V_{\rm max}$ function and the simulated systems. The size of the measured $V_{\rm max}$ function deviation relative to the size of the scatter will tell us about how readily the Milky Way $V_{\rm max}$ function is realized in each of our models.

The first step is to define the variation within the set of $V_{\rm max}$ for a given halo mass-dark matter model–galaxy formation model combination. The n most massive satellites of the ith simulated halo are selected, and their values of $V_{\rm max}$ are sorted into descending order. The $V_{\rm max}$ distribution is then $\{V_{i,1}, V_{i,2}, V_{i,3} \dots V_{i,n}\}$; here we have omitted the 'max' subscript for clarity. We can define the difference between this ith halo distribution with respect to the jth halo distribution thus:

$$Q_{i,j} = \frac{\sum_{k=1}^{n} |V_{i,k} - V_{j,k}|}{\sum_{k=1}^{n} (V_{i,k} + V_{j,k})}, \tag{7}$$

and if there are N realizations of the model in question, the mean Q for the ith distribution, $\bar{Q}_i$, is

$$\bar{Q}_i = \frac{1}{N-1} \sum_{j \neq i} Q_{i,j}. \tag{8}$$

Similarly, if we substitute the ith simulated $V_{\rm max}$ distribution to instead be $\{V_{\rm MW,1}, V_{\rm MW,2}, V_{\rm MW,3} \dots V_{\rm MW,n}\}$, i.e. the observed $V_{\rm max}$ distribution of the Milky Way satellites, then we obtain $Q_{\rm MW}$:

$$Q_{\rm MW} = \frac{1}{N} \sum_{j} Q_{\rm MW,j}. \tag{9}$$

The probability that a $V_{\rm max}$ function with $Q_{\rm MW}$ could be drawn from the parent distribution is then $P(> Q_{MW})$, where P is the cumulative distribution of $\bar{Q}$.

We expand on the method described above to describe how we select satellites. The luminosities calculated for the satellites enable us to take account of the fact that the brightest satellites, for which the velocity dispersions have been measured with the highest precision, need not necessarily reside in the most massive haloes. Therefore, we consider two options for selecting our top 'n' satellites to be matched to observations: (i) select the n brightest V-band satellites and (ii) select the n highest $V_{\rm max}$ satellites. We compare the results from these two approaches in Section 3.1.3.

2.5 Sterile neutrino matter power spectra

In addition to CDM, we consider keV-scale, resonantly produced sterile neutrino dark matter. The latter constitutes part of a larger particle physics model called the neutrino minimal standard model (νMSM), which explains neutrino oscillations and baryogenesis in addition to yielding a dark matter candidate, see Boyarsky, Ruchayskiy & Shaposhnikov (2009) for a review. The keV sterile neutrino behaves like WDM, in that it free streams out of small perturbations in the early Universe. The resulting matter power spectrum cutoff is influenced by two parameters: the sterile neutrino mass, $M_{\rm s}$, and the lepton asymmetry in which the dark matter is generated (Shi & Fuller 1999; Laine & Shaposhnikov 2008; Boyarsky et al. 2009; Ghiglieri & Laine 2015; Venumadhav et al. 2016). We parametrize the lepton asymmetry as L_6, which is defined as $10^6 \times$ the difference in lepton and antilepton abundance normalized by the entropy density. The power spectrum cutoff shifts to smaller scales for larger values of the mass, as is the case for thermal relic WDM. By contrast, the behaviour with lepton asymmetry is non-monotonic; for a recent discussion see Lovell et al. (2016b).

We focus on the parameter space that is roughly in agreement with the recent observations of the 3.5 keV emission line detected in Bulbul et al. (2014); Boyarsky et al. (2014, 2015), which requires a sterile neutrino mass of 7 keV and a lepton asymmetry in the range $L_6 = [8, 11.2]$, where the uncertainty in L_6 is dominated by the uncertainty in the dark matter content of the target galaxies and galaxy clusters. The recent study by Ruchayskiy et al. (2016) set a more stringent lower limit of $L_6 > 9$; however, $L_6 = 8$ remains of interest as it has the shortest free-streaming length obtainable by a 7 keV sterile neutrino of any lepton asymmetry. We therefore select primarily three models for our study, $L_6 = 8, 10, 12$, in order to span the range of L_6 that is in agreement with the detected decay line. From hereon in we refer to these models as LA8, LA10 and LA12. We also briefly consider four further models to probe a larger range of free-streaming lengths: three 7 keV particles (L_6=[14,18,120]) and one 10 keV sterile neutrino with $L_6 = 7$.

We first calculate the momentum distribution functions for our three sterile neutrino models using the methods and code of Laine & Shaposhnikov (2008) and Ghiglieri & Laine (2015). From these distribution functions we then derive the matter power spectra by means of a modified version of the CAMB Boltzmann-solver code (Lewis, Challinor & Lasenby 2000). The results are plotted in Fig. 2 as dimensionless matter power spectra. All three models exhibit a cutoff, and the cutoff position shifts to larger scales – smaller wavenumbers – with increasing L_6.[2]

Also plotted is the power spectrum of the 2.3 keV thermal relic studied by Wang et al. (2016), who showed, using N-body simulations, that, since halo concentrations are lower for WDM than for CDM haloes, this particular model required subhaloes of $V_{\rm max} \sim 1.17$ times higher than ΛCDM to fit the kinematics and photometry of Fornax. We will use this correction factor in our study to illustrate the impact of lower sterile neutrino halo densities on their hosted galaxies. We caution that this factor was derived for only one satellite and for a dark matter model that has a larger

[2] $L_6 = 8$ is the model for which the cutoff is located at the smallest scale, since for smaller L_6 the influence of resonant production is weaker and thus the cutoff moves to larger scales.

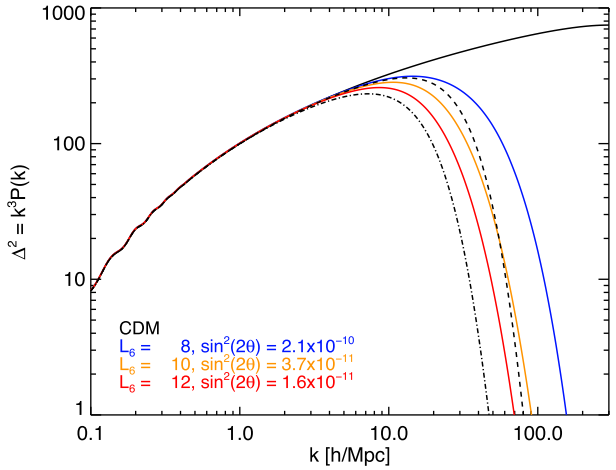

Figure 2. Matter power spectra for our four dark matter models: CDM (black, solid), LA8 (blue), LA10 (orange) and LA12 (red). The black dot–dashed line denotes the power spectrum of the 2.3 keV thermal relic studied in Wang et al. (2016), and the dashed line is the 3.3 keV thermal relic power spectrum used in COCO-WDM (Bose et al. 2016).

free-streaming length than any of our three primary WDM models. Our results for WDM should therefore be considered as a rough approximation, rather than rigorous predictions. In addition, central halo masses $<1.4 \times 10^{12}\,\mathrm{M}_{\odot}$ are disfavoured for these models in the current model of reionization feedback by virtue of their low satellite counts (Lovell et al. 2016b); however, we include them here for completeness.

The application of the GALFORM feedback model is complicated in WDM-style models by the dependence of the feedback strength on the halo circular velocity. In GALFORM, the strength of feedback is modelled as a power law of the circular velocity, where the power-law index is denoted γ. The lower circular velocities of WDM haloes lead to the result that WDM models run using the CDM model parameters underpredict the number of galaxies with $M_V < -16$. A discussion of this issue can be found in Kennedy et al. (2014) and Lovell et al. (2016b). We recalibrate the model against the b_J band luminosity function and find that $\gamma = 3.15$ is a good fit to the observational data for all three of our sterile neutrino models as opposed to $\gamma = 3.4$ for the standard CDM model. We therefore adopt $\gamma = 3.15$ for LA8, LA10 and LA12, and retain $\gamma = 3.4$ for CDM.

We also make the following assumptions with regard to the stripping algorithm in the sterile neutrino models:

(i) Given that WDM subhaloes deviate slightly from NFW profiles (Colín, Valenzuela & Avila-Reese 2008; Lovell et al. 2014; Ludlow et al. 2016), a complete study would re-evaluate whether the V_{max}–V_{vir} relation (equation 6) would need to be recalibrated. For simplicity we use equation (6) to calculate V_{max} for all of our models.

(ii) Hydrodynamical models of WDM have shown that WDM subhaloes exhibit the same degree of mass-loss as CDM haloes (Lovell et al. 2016a), thus equation (2) is equally valid for our sterile neutrino simulations.

The stripping model is calibrated to CDM simulations, in which the halo mass–concentration relation will play a key role in the stripping rates. This relation changes for, and between, different WDM models. Therefore, a precise prediction for the $z = 0$ V_{max} functions for a given WDM model requires that we calibrate each model

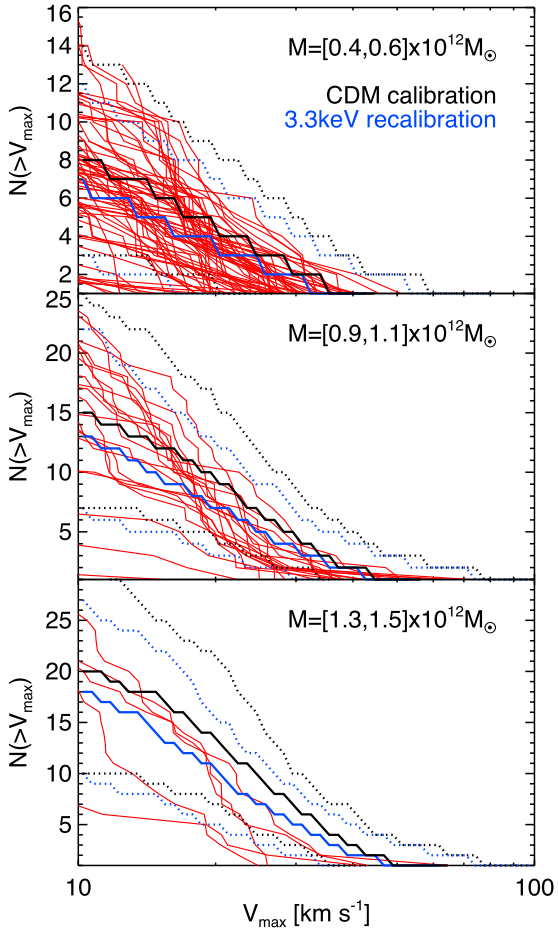

Figure 3. The V_{max} functions of 3.3 keV thermal relics as predicted by the COCO-WDM simulation and the PCH + stripping method. We include the PCH data as computed using the CDM calibration ($\bar{A} = 1.4$) in black and the recalibration for the 3.3 keV relic ($\bar{A} = 1.1$) in blue. The COCO-WDM V_{max} functions are shown as red lines.

to an N-body simulation of that specific model. We do not have N-body simulations for any of the sterile neutrino models discussed below; instead we make a qualitative prediction for how our results would change by calibrating our model to that of a 3.3 keV relic as used in the COCO-WDM simulation (Bose et al. 2016), which is a good approximation to our LA8 model. We repeat the same process discussed above as applied to COCO-CDM, with the CDM matter power spectrum replaced by that of a 3.3 keV thermal relic, both with the CDM-calibration value $\bar{A} = 1.4$ and a recalibrated version with $\bar{A} = 1.1$. We present our results in Fig. 3.

The original calibration works well for the lowest mass halo bin, but systematically overpredicts the V_{max} functions of the 1.0×10^{12} and $1.4 \times 10^{12}\,\mathrm{M}_{\odot}$. This is because the WDM haloes are less concentrated than the CDM and thus the stripping rates are higher. Our recalibration ameliorates some of the discrepancy, although it still overpredicts the V_{max} functions of the two more massive haloes, in order to not underpredict the $0.5 \times 10^{12}\,\mathrm{M}_{\odot}$ mass functions. The mean suppression of the recalibrated model relative to the original at a V_{max} of 20 km s^{-1}of 30 per cent, even for this relatively warm model, and is therefore significant. We adopt $\bar{A} = 1.1$ for all of our sterile neutrino models, and state how the results would change for a precise calibration to each separate model where appropriate.

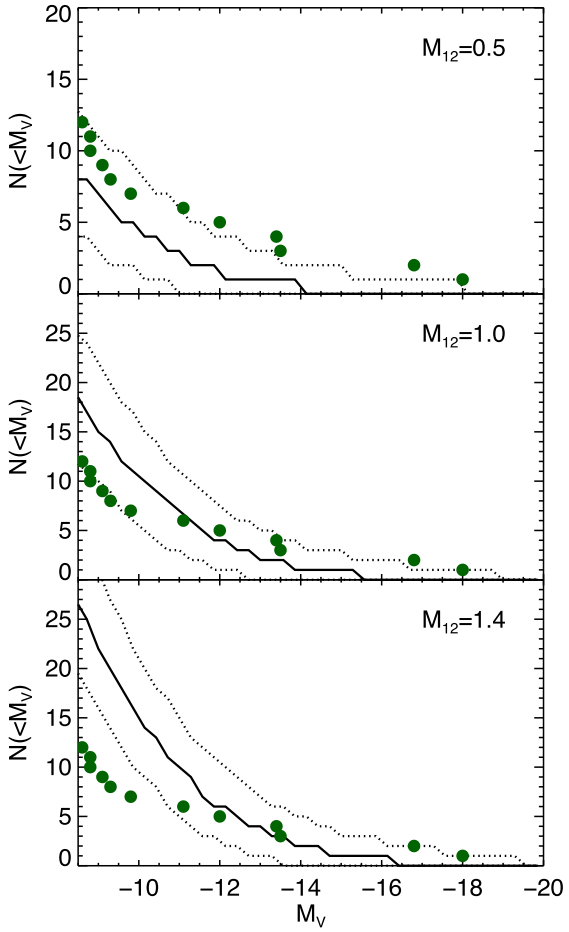

Figure 4. Cumulative satellite luminosity function for the LC16 ΛCDM galaxy formation model and three halo masses. The solid lines denote the median number of satellites brighter than M_V across all the realizations and the dotted curves mark the 5 and 95 percentiles. The top, middle and bottom panels show the mass functions for central haloes of mass 0.5, 1.0 and $1.4 \times 10^{12}\,\mathrm{M}_\odot$ respectively. The circles mark the observed Milky Way satellite luminosity function.

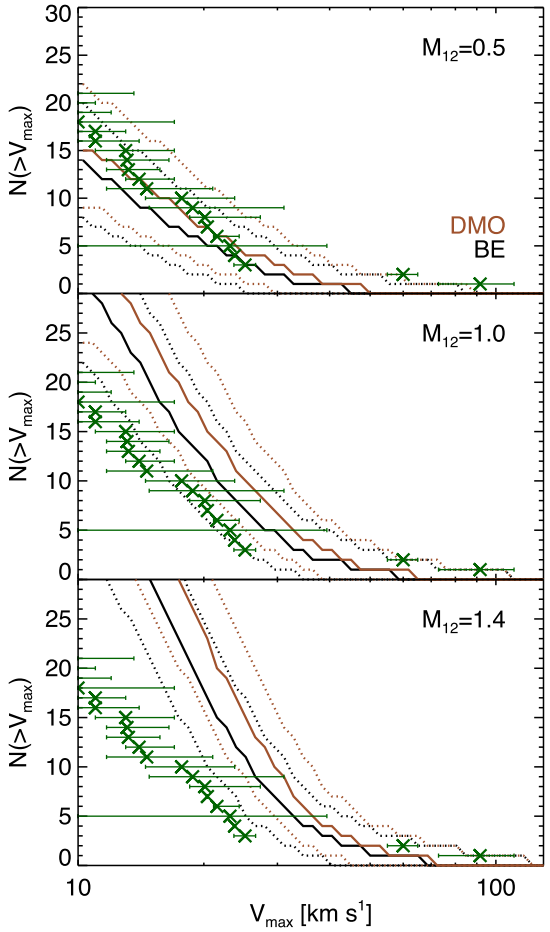

Figure 5. Cumulative satellite $V_{\rm max}$ function for the ΛCDM-LC16 model when the correction for baryonic effects is applied (black) and not (brown). The top, middle and bottom panels show the mass functions for central haloes of mass 0.5, 1.0 and $1.4 \times 10^{12}\,\mathrm{M}_\odot$ respectively. The inferred $V_{\rm max}$ function of the Milky Way satellites is shown as green crosses. Note that error bars are not included for two of the satellites, Fornax and Draco, because these were not calculated by Sawala et al. (2016b).

3 RESULTS

In this section we show to what degree our models agree with the observed luminosities and $V_{\rm max}$ of the Milky Way satellites, and then analyse the Too Big To Fail problem using the statistic developed by Jiang & van den Bosch (2015). We first consider the effect of baryon physics on the CDM $V_{\rm max}$ function in Section 3.1, and then apply our preferred baryon model to the sterile neutrino models in Section 3.2.

3.1 Baryon physics

3.1.1 Luminosity functions

We begin our discussion of the results with the luminosity functions of each of our GALFORM models. In Fig. 4, we present the luminosity functions for the LC16 model and three halo masses ([0.5, 1.0, 1.4] $\times 10^{12}\,\mathrm{M}_\odot$). We also include the observed luminosity function of the Milky Way satellites, which we assume to be complete for the range of luminosities considered.

The most striking difference between the observations and all four models is the steepness of all the simulated luminosity functions relative to that of the Milky Way. This is realized as a dearth of large and small Magellanic Cloud (LMC and SMC) candidates for the $5 \times 10^{11}\,\mathrm{M}_\odot$ halo and an overproduction of bright satellites for the central mass of $1.4 \times 10^{12}\,\mathrm{M}_\odot$. However, the $1 \times 10^{12}\,\mathrm{M}_\odot$ returns a reasonable match to the observations.

3.1.2 $\mathrm{V}_{\rm max}$ functions for luminous satellites

Identifying which satellites are luminous enables us to make a more accurate comparison between the simulated $V_{\rm max}$ function and that inferred for the Milky Way satellites. The $V_{\rm max}$ function is influenced by early loss of baryons from a halo, as described by Sawala et al. (2016b). We illustrate the importance of this effect in Fig. 5, in which we show the median cumulative $V_{\rm max}$ functions for CDM in two cases, with the baryon suppression of equation (2) turned off (brown lines) and turned on (black lines). Unlike in Fig. 1, we only plot satellites that are luminous.

We first discuss the case in which the suppression of $V_{\rm max}$ by baryon effects is not taken into account. For the lowest mass halo, the ΛCDM model provides a good description of the data, except for the lack of any LMC counterparts. For a halo mass of $1.0 \times 10^{12}\,\mathrm{M}_\odot$, the model tends to overpredict the observed $V_{\rm max}$ function, although the uncertainties in $V_{\rm max}$ are large enough for the model to be

consistent with the data. For a halo mass of $1.4 \times 10^{12}\,M_\odot$, the discrepancy is large enough that it cannot be explained by uncertainties in the V_{max} measurements. Applying the correction to V_{max} due to baryon effects, as described by equation (2), produces a significant shift in the predicted V_{max} function. Now the models with halo masses of 0.5 and $1 \times 10^{12}\,M_\odot$ are entirely consistent with the data but the model with the largest halo mass is still ruled out. We therefore conclude that the suppression of satellite mass caused by the early loss of baryons from the halo is crucial in order to explain the observed V_{max} function, in agreement with Sawala et al. (2016b) and Fattahi et al. (2016), but with the added constraint that the mass of the Milky Way halo should be lower that about $1.4 \times 10^{12}\,M_\odot$.

3.1.3 Statistical comparison of simulated and observed V_{max} *functions*

The strength of our PCH method, as compared to hydrodynamical simulations like Sawala et al. (2016b) and Fattahi et al. (2016), is that it is practical to run hundreds of merger trees very quickly and thus build good statistical samples. We can therefore calculate what proportion of simulated systems returns a V_{max} function that is a good match to that of the Milky Way satellites, and thus quantify the quality of the agreement between observations and the model V_{max} functions shown in Fig. 5. This is done by extracting the n most massive luminous satellites and calculating the Q statistic for this distribution using the methods of Jiang & van den Bosch (2015) as summarized in Section 2.4. We compare the value of Q obtained for the Milky Way system with respect to the PCH results, denoted Q_{MW}, to the distribution of PCH Q. The closer Q_{MW} is to the centre of the Q distribution, the better the agreement is between the model and the observations.

In Fig. 6, we plot the distributions of Q for the 0.5×10^{12}, 1.0×10^{12} and $1.4 \times 10^{12}\,M_\odot$ haloes and four algorithms for generating V_{max} functions. These algorithms are

(i) All satellites, baryon effects not applied (also referred to as 'DMO').
(ii) All satellites, baryon effects included (BE).
(iii) Satellites ordered by luminosity, baryon effects not applied (Lum).
(iv) Satellites ordered by luminosity, baryon effects included (Lum+BE).

The number of satellites selected in each case is $n = 10$.

For all three halo masses, we measure an important effect on the Q distribution between the different algorithms. The application of the feedback suppression factor increases the scatter slightly between distributions relative to the base model [model (i) above] and thus translates each curve to the right by 0.01 units in Q irrespective of the halo mass. A marginally larger shift occurs when satellites are first sorted by luminosity, and the two effects combined produce a shift of 0.02 Q relative to the base.

There is also a trend on the value of Q_{MW}. When considering the haloes of mass 1.0×10^{12} and $1.4 \times 10^{12}\,M_\odot$, luminosity ordering lowers Q_{MW} as the greater scatter grows closer to encompassing the observational data. The baryonic effects produce a stronger effect in the same direction because the increase in the *scatter* is accompanied by a fall in the *mean* V_{max} function, and thus closer to the Milky Way satellite V_{max} function as shown in Fig. 5. The application of these two lower Q_{MW} still further, by a total of 0.04 points relative to the base model. In combination with the greater scatter within the simulated distributions, the overlap between Q_{MW} and the Q distributions improves significantly. Halo masses that would have been incompatible with the Milky Way satellite V_{max} function under the base model are now very possible, if still rare. Note that this improvement does not occur for the lightest halo mass; however, the base model V_{max} function is itself in good agreement with that of the Milky Way satellites, so further suppression results in stronger disagreement with the data.

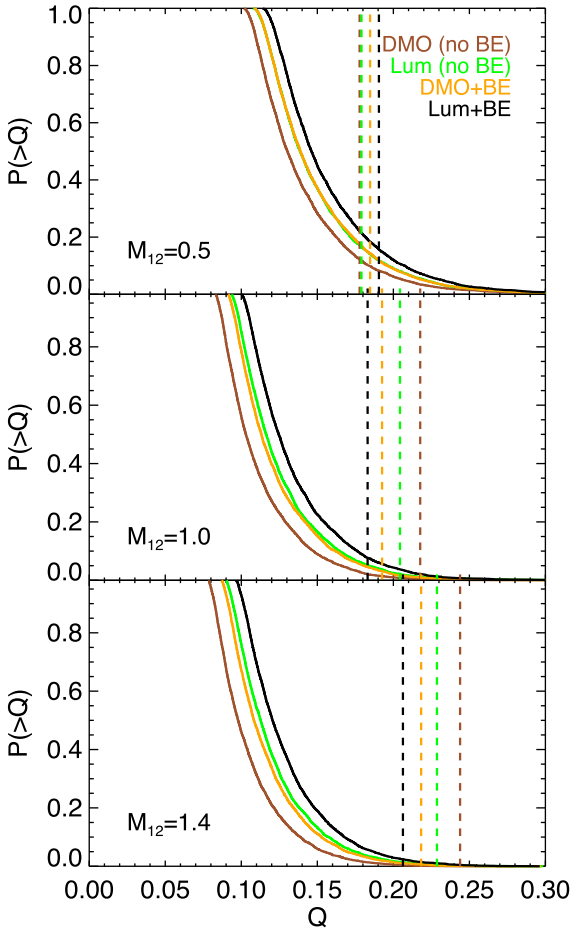

Figure 6. Cumulative Q statistic function for the CDM-LC16 model using our four V_{max} variations: luminous satellites (brown), baronic effects applied (orange), luminosity-ordered satellites (green) and luminosity ordered with baryonic effects applied (black). Solid lines denote the cumulative Q statistic functions, and dashed lines the corresponding value of Q_{MW}. The top, middle and bottom panels show the mass functions for central haloes of mass 0.5, 1.0 and $1.4 \times 10^{12}\,M_\odot$ respectively.

3.1.4 Probability of drawing the Milky Way satellite V_{max} *function from simulated* V_{max} *distributions*

In the previous section, we showed that the mean V_{max} function amplitude correlates with central halo mass, such that for a given halo selection algorithm there is a 'sweet spot' halo mass at which the probability of drawing a Milky Way like satellite V_{max} function is maximized. The probability that the Milky Way distribution can be drawn from a V_{max} distribution at fixed host halo mass is quantified by the cumulative probability distribution $P(>Q_{MW})$. If $P(>Q_{MW}) \ll 0.01$ then that halo mass-model combination is ruled out. Therefore, we calculate $P(>Q_{MW})$ as a function of halo mass for our set of 14 central halo masses and plot the results for our four V_{max} function algorithms in Fig. 7. Note that in all four cases we

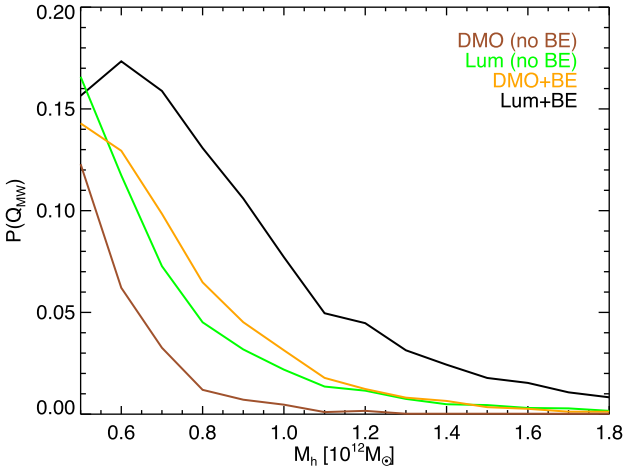

Figure 7. The probability that a Milky Way like satellite $V_{\rm max}$ distribution is drawn from the simulated distributions as a function of halo mass. Our varieties of $V_{\rm max}$ functions are shown using the same colours as in Fig. 6: luminous satellites (brown), luminosity-ordered satellites (green), baryonic effects applied (orange) and luminosity ordered with baryonic effects applied (black).

select 10 satellites, the difference is solely in how they are selected and processed.

All four curves show preferences for lighter haloes; however, the luminosity-ordered + feedback suppression shows a shift towards higher mass haloes. The amplitude of the curves is lowest for the base model, which registers an effective zero probability for haloes more massive than $1.3 \times 10^{12}\,{\rm M}_{\odot}$. Luminosity ordering increases the probability across all halo masses, feedback suppression further still and the highest probabilities are found for the luminosity-ordering + feedback suppression algorithm. In this case, even halo masses of $1.8 \times 10^{12}\,{\rm M}_{\odot}$ can host Milky Way like satellite $V_{\rm max}$ functions, albeit very rarely. One should also note that we showed in Fig. 1 that our stripping model overpredicts the number of $\sim$25 km s^{-1}subhaloes; therefore, a more accurate stripping model will return probabilities higher than those calculated here.

3.2 Sterile neutrino dark matter

We now consider the changes that would be made to our results if the dark matter were a WDM candidate, specifically the sterile neutrino. In Fig. 8, we plot the luminosity functions of three sterile neutrino models, LA8, LA10 and LA12, in addition to CDM. The luminosity functions between CDM and LA8 are remarkably similar, which is in part due to our use of weaker, recalibrated feedback. The number of satellites is suppressed in the other two models; however, not enough to achieve agreement with the observations for the highest mass halo. Any comprehensive and accurate model of galaxy formation would therefore still require stronger feedback in low-mass galaxies than that used here, although the adoption of WDM may play a subdominant part in achieving the necessary agreement.

Having shown that the sterile neutrino models produce acceptable numbers of satellites, we now consider their $V_{\rm max}$ functions. We apply the suppression factor from baryon effects from equation (2) to all four dark matter models and plot the results in Fig. 9. There is a systematic decrease of the $V_{\rm max}$ function with free-streaming length, to the extent that LA12 hosts $\sim$50 per cent fewer satellites with $V_{\rm max} > 10\,{\rm km\,s^{-1}}$than ΛCDM. This suppression moves the sterile neutrino $V_{\rm max}$ functions closer to the measured Milky

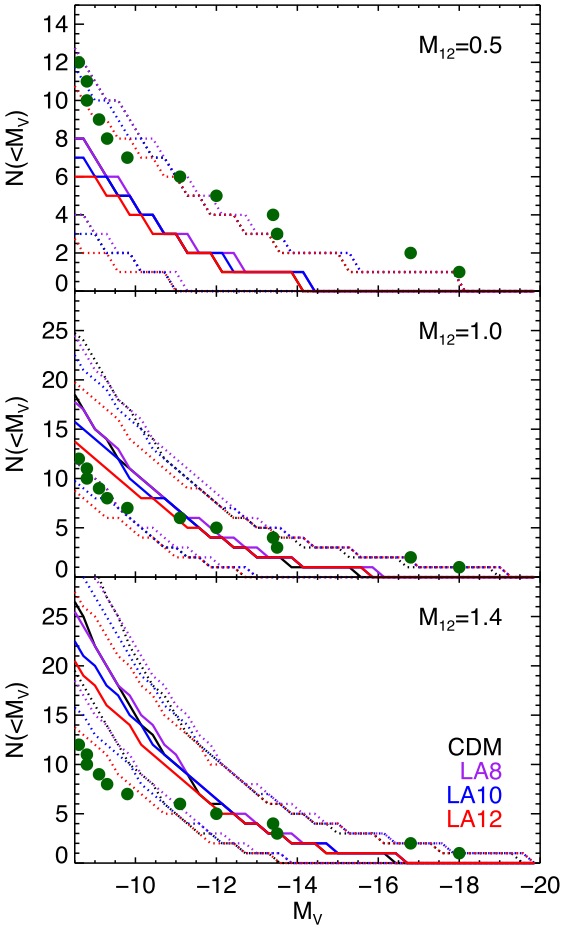

Figure 8. Cumulative satellite luminosity function for four dark matter models and three halo masses. The galaxy formation model is LC16, with a refitted $\gamma_{\rm SN}$ parameter for the sterile neutrino models. The solid lines denote the median number of satellites brighter than M_V across all of our realizations, and the dotted curves mark the 5 and 95 percentiles. Each dark matter model is denoted by a different colour: CDM (black), LA8 (purple), LA10 (blue) and LA12 (red), as indicated in the legend. The top, middle and bottom panels show the mass functions for central haloes of mass 0.5, 1.0 and $1.4 \times 10^{12}\,{\rm M}_{\odot}$ respectively. The circles mark the observed Milky Way satellite luminosity function.

Way satellite $V_{\rm max}$ function. The improvement is even greater for the $1.4 \times 10^{12}\,{\rm M}_{\odot}$ halo when we take into account the different concentration–mass relation of WDM models, as parametrized by our dwarf spheroidal $V_{\rm max}$ correction value of 1.17 ; for the 1.0×10^{12} and $0.5 \times 10^{12}\,{\rm M}_{\odot}$ haloes the agreement with the modified $V_{\rm max}$ function is instead weaker, since the theoretical $V_{\rm max}$ functions are now oversuppressed. Thus, in general the suppressed $V_{\rm max}$ functions and lower concentrations of the sterile neutrino models combine to give better agreement with the observations at larger halo masses than in ΛCDM.

To end this section, we calculate the probability of drawing Milky Way like satellite $V_{\rm max}$ functions from our sterile neutrino $V_{\rm max}$ distributions, once again using the Q distribution-$Q_{\rm MW}$ combination from Section 3.1.4. We present our results as a function of host halo mass in Fig. 10. When we assume the same values of $V_{\rm max}$ for the Milky Way satellites in the sterile neutrino models as in CDM, we find that the amplitude of the probability curve remains roughly the same. The difference instead comes from a shift to larger masses of the probability distribution peak, which reflects how the decrease

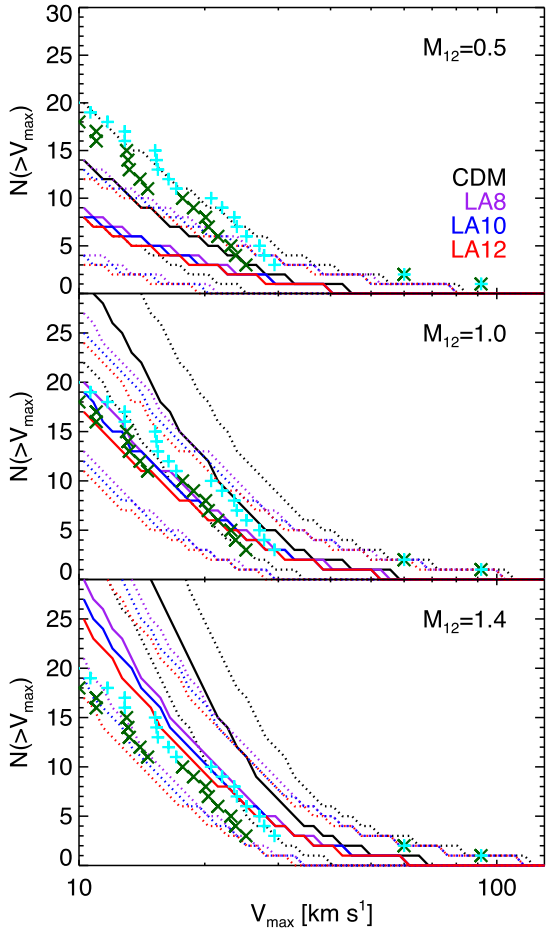

Figure 9. Cumulative satellite $V_{\rm max}$ function for the same dark matter models and halo masses presented in Fig. 8. We include all luminous satellites, and have applied the baryonic feedback correction from equation (2). The solid lines denote the median $V_{\rm max}$ across all of our realizations, and the dotted curves again mark the 5 and 95 percentiles. The colour-dark matter model correspondence is the same as in Fig. 8: CDM (black), LA8 (purple), LA10 (blue) and LA12 (red). The top, middle and bottom panels again show the mass functions for central haloes of mass 0.5, 1.0 and 1.4 $\times$ 10^{12} $M_\odot$ respectively. The dark green crosses mark the inferred Milky Way satellite $V_{\rm max}$ function assuming CDM. We also include cyan plus signs, for which the $V_{\rm max}$ values of the dwarf spheroidals (but not the Magellanic Clouds) are multiplied by the factor of 1.17 suggested by the results of Wang et al. (2016). We have not attempted to correct for incompleteness in the observed satellite sample. Therefore, these values constitute a lower bound on the expected Milky Way satellite $V_{\rm max}$ function. The $V_{\rm max}$ error bars have been omitted for clarity.

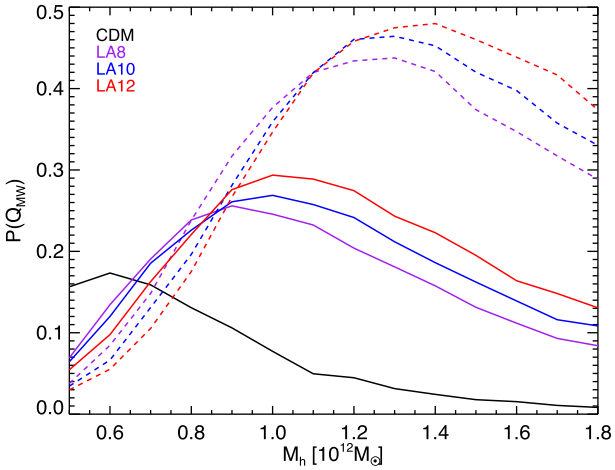

Figure 10. The probability that a Milky Way like satellite $V_{\rm max}$ distribution is drawn from the simulated distributions as a function of halo mass. The $V_{\rm max}$ function selection is made using the luminosity-ordered + baryonic effects correction scheme, and the galaxy formation model is LC16 with recalibration for the sterile neutrino models. The colour-dark matter model correspondence is the same as in Fig. 8: CDM (black), LA8 (purple), LA10 (blue) and LA12 (red). Solid lines denote results calculated when the observed values of $V_{\rm max}$ are derived from CDM simulations, and dashed where the Wang et al. (2016) factor is applied.

in the number of satellites requires a more massive host halo to hit the observed target. The consequences at the largest halo masses are significant: a fit to the 1.0 $\times$ 10^{12} $M_\odot$ halo is over three times as likely in LA12 than it is in CDM, and a fit to the 1.4 $\times$ 10^{12} $M_\odot$ halo eight times more likely.

More impressive still is the contribution made by the lower concentrations. The adoption of the Wang et al. (2016) correction factor improves the probability by over a factor of 2 as compared to the CDM–$V_{\rm max}$ values, with the peak in the probability distribution located as high as 1.4 $\times$ 10^{12} $M_\odot$. This result reflects the fact that the observed $V_{\rm max}$ function has not only a higher amplitude in WDM, which can be achieved just by choosing a larger halo, but is also steeper, and therefore has a shape more in keeping with that of the simulated data. We stress, however, that this result is purely illustrative because it is based on just one WDM model (a 2.2 keV thermal relic) and one observed satellite (Fornax), therefore fits of many more satellites to many more dark matter models are required to ascertain the precise boost to the probability provided by lower concentration haloes. We also note that the stripping method has been calibrated to just one WDM model, the 3.3 keV thermal relic. This model is similar to our least extreme model, LA8, and may not be appropriate for the other two models. We expect that these models will experience even more stripping than we predict here, pushing the preferred halo mass still higher.

We end this section with a study of the probability of hosting a Milky Way like $V_{\rm max}$ function as a function of the dark matter power spectrum cutoff. We parametrize each of our models using the position of the peak of each matter power spectrum, which we denote $k_{\rm peak}$. For CDM this value is formally infinite, therefore we consider the inverse of the peak, $k_{\rm peak}^{-1}$. We consider three halo masses (0.5, 1.0, 1.4 $\times$ 10^{12} $M_\odot$) and six sterile neutrino models (7 keV, L_6=120, 18, 14, 10 and 8, plus 10 keV, L_6 = 7), and plot the results in Fig. 11. In order to make the connection to particle physics experiments and previous work on the subject, we also include equivalent thermal relic masses for our models on the top x-axis. These are the thermal relic masses that have the same value of $k_{\rm peak}$ as our sterile neutrino models, with their matter power spectra calculated using the procedure of Viel et al. (2005).

The value of the probability, $P(>Q_{\rm MW})$, correlates with $k_{\rm peak}^{-1}$ for all three halo masses. For the two more massive haloes, the trend is positive as a reflection of the suppression of the $V_{\rm max}$ function with $k_{\rm peak}^{-1}$, for the lightest halo the trend is reversed due to oversuppression. The probability may increase by as much as a factor of 3 when the Wang et al. (2016) factor is applied. However, we reiterate that this correction is based on just one WDM model and one satellite. A precise prediction will require a fit for every satellite with every model of interest, which we defer to later work.

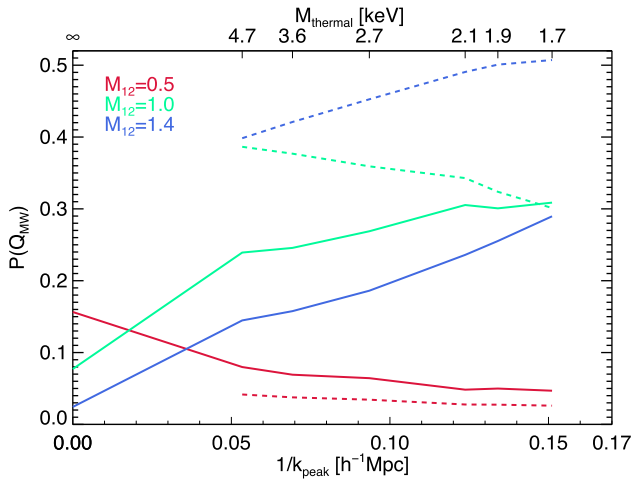

Figure 11. The probability that a Milky Way like satellite $V_{\rm max}$ distribution is drawn from the simulated distributions as a function of the inverse of the dimensionless matter power spectrum peak wavenumber, $k_{\rm peak}^{-1}$. The $V_{\rm max}$ function selection is made using the luminosity-ordered + baryonic effects correction scheme, and the galaxy formation model is LC16 with recalibration for the sterile neutrino models. The thermal relic masses corresponding to the value of $k_{\rm peak}^{-1}$ for each of our models are displayed on the top axis. The values 1.7, 1.9, 2.1, 2.7 and 3.6 keV correspond to the 7 keV sterile neutrino with L_6=120, 18, 14, 10 and 8 respectively; the model at $M_{\rm thermal}$ =4.7 keV is a 10 keV sterile neutrino with $L_6 = 7$. We do not include $L_6 = 12$ (LA12) in this plot due to a lack of space. The colours correspond to different host halo masses $0.5 \times 10^{12}\,{\rm M}_\odot$ (red), $1.0 \times 10^{12}\,{\rm M}_\odot$ (green) and $1.4 \times 10^{12}\,{\rm M}_\odot$ (blue). Solid lines denote results calculated when the observed values of $V_{\rm max}$ are derived from CDM simulations, and dashed where the Wang et al. (2016) factor is applied.

4 CONCLUSIONS

The central densities of satellites have been the subject of much recent study. Observations have been used to estimate the masses of satellite galaxies and simulations have improved sufficiently to make robust predictions for satellite density profiles. The observations were found by Boylan-Kolchin et al. (2011, 2012) to be discrepant with *N*-body (dark matter only) simulations, which overpredict the inner densities measured for the brightest Milky Way satellites. This issue became known as the 'too big to fail' problem.

Many solutions have been suggested, and in some cases they complement one another. These include: assuming a relatively low mass for the halo of the Milky Way (Wang et al. 2012; Cautun et al. 2014); changing the cosmological parameters (Polisensky & Ricotti 2014); the creation of a central core by supernova feedback (Brooks & Zolotov 2014); a reduction in the value of $V_{\rm max}$, reflecting lower halo growth induced by early mass-loss (Sawala et al. 2016b); assuming that the dark matter is self-interacting (Zavala et al. 2013), that it couples to radiation (Schewtschenko et al. 2016), or that it consists of sterile neutrinos (Lovell et al. 2012), in which cases satellite haloes are less dense that in the standard CDM model.

In this study, we considered three of these possible solutions, namely a low Milky Way halo mass, the suppression of $V_{\rm max}$ by baryonic effects and sterile neutrino dark matter. In addition, we considered the impact of selecting satellites by stellar mass or luminosity rather than by halo mass, as is done in an *N*-body simulation. Each possibility was considered separately and in concert in order to establish which combination of factors would provide the best match to the measured Milky Way satellite $V_{\rm max}$ function.

We computed Milky Way luminosity and $V_{\rm max}$ functions for 14 Milky Way halo masses in the range $[0.5, 1.8] \times 10^{12}\,{\rm M}_\odot$ using a modification of the Lacey et al. (2016) version of the GALFORM semi-analytic galaxy formation model, described in Lacey et al. (2016), that was adapted to be run assuming an underlying Planck cosmology, PCH halo merger trees and the subhalo stripping algorithm introduced by Jiang & van den Bosch (2015). The dark matter subhaloes were populated with galaxies by GALFORM and we calculated the suppression of $V_{\rm max}$ by baryonic effects using the parametrization introduced by Sawala et al. (2016b). We recalibrated the semi-analytic stripping model against the CDM-COCO simulation, and recovered a good match between the PCH and *N*-body $V_{\rm max}$ functions for $V_{\rm max} \geq 20\,{\rm km\,s^{-1}}$.

The sterile neutrino model was a 7 keV mass particle, chosen to be consistent with the decay interpretation of the otherwise unexplained 3.55 keV line signal detected in clusters of galaxies and in M31 (Boyarsky et al. 2014; Bulbul et al. 2014).The measured flux from these targets implies a mixing angle for the sterile neutrino in the range $\sin^2(2\theta) = [2, 20] \times 10^{-11}$. This corresponds to a generation lepton asymmetry approximately in the range L_6=[8,12]. The value of the lepton asymmetry plays a role in setting the free-streaming length; therefore we adopted three values of L_6: 8, 10 and 12. $L_6 = 8$ has the shortest free-streaming length and $L_6 = 12$ the longest of the models we consider. For each combination of these three sterile neutrino models, and for CDM, with the chosen halo masses we generated 5000 merger trees in order to take account of the stochastic scatter introduced by different merger histories.

We showed that the models predict luminosity functions that tend to be steeper than, but still consistent with the data, even in the luminosity range in which the satellite census is thought to be complete (Fig. 4). Models that predict the correct number of $M_V = -10$ galaxies produce LMC-like satellites in less than 10 per cent of realizations. The suppression at low luminosities in the sterile neutrino models leads to even better agreement with the observed luminosity function.

A similar pattern was found in the $V_{\rm max}$ functions, in that models that host Magellanic Cloud analogues tend to overpredict the number of less massive satellites unless they have a rather small total mass (Fig. 5). As found by Sawala et al. (2016a), this tension is eliminated when the suppression of $V_{\rm max}$ by baryonic effects, which decreases the median number of Milky Way satellites with $V_{\rm max} > 20\,{\rm km\,s^{-1}}$ from 16 to 12, is taken into account. The agreement with observations is better still for the sterile neutrino models, especially since the lower concentrations of sterile neutrino haloes translate into a lower host halo $V_{\rm max}$.

In order to determine how likely the Milky Way $V_{\rm max}$ function is to have been drawn from our PCH-generated $V_{\rm max}$ distributions, we characterize the variation between individual halo realizations using the Q statistic introduced by Jiang & van den Bosch (2015). The probability that the Milky Way satellite $V_{\rm max}$ function could be drawn from that distribution is then $P(>Q_{\rm MW})$, where $Q_{\rm MW}$ is the value of Q for the Milky Way satellite $V_{\rm max}$ function relative to the simulated version. We find that, for halo masses $\geq 1 \times 10^{12}\,{\rm M}_\odot$, the selection of the brightest subhaloes rather than all luminous subhaloes can increase $P(>Q_{\rm MW})$ by a factor of 10, and the correction of $V_{\rm max}$ due to for baryonic effects by up to a further factor of 2 (Fig. 7). Sterile neutrino models have a higher likelihood than CDM models, and $P(>Q_{\rm MW})$ is correlated with the free-streaming length. This trend is reversed for smaller halo masses, due to the lack of massive satellites in the sterile neutrino models.

We have thus shown that satellite $V_{\rm max}$ functions like that of the Milky Way are generated in the CDM cosmology. They are more

common in sterile neutrino cosmologies, which allow for satellites to reside in more massive haloes, of which there are fewer still than in CDM. This model also has the attraction that it matches the data for a set of sterile neutrino parameters that account for the 3.5 keV line feature.

There are many uncertainties that remain in the sterile neutrino calculation presented in this paper. The first is that the exact degree of halo tidal stripping is unknown; in principle this needs to be assessed using simulations for each sterile neutrino case (Bozek et al. 2016). It also remains to be seen whether these sterile neutrino models generate enough faint ($M_V > -8$) satellites (Lovell et al. 2016b; Schneider 2016) or match the Lyman-α forest flux measurements (Viel et al. 2013; Schneider et al. 2014; but see also Garzilli, Boyarsky & Ruchayskiy 2015). There is currently enough uncertainty in both the galaxy formation model and the observational constraints that these models cannot be ruled out; however, tighter constraints from future observational surveys may help establish if these models are viable.

With respect to the CDM cosmology, we find that Milky Way like systems are rare but by no means impossible. This represents a refinement on other studies, such as Sawala et al. (2016b) and Fattahi et al. (2016), that similarly find the Milky Way systems in high-resolution simulations typically underpredict the satellite $V_{\rm max}$ functions but not to a severe degree. More stringent tests of the model will be realized as both observations and simulations attain the ability to measure the velocity dispersions of faint galaxies: if the census of the 20 brightest satellites is already complete then the disagreement with observations is very strong (Jiang & van den Bosch 2015). Hints in this direction have been discovered in a new set of fairly bright but extended satellites of the Milky Way (Caldwell et al. 2016) and M31 (Collins et al. 2013) in which the measured velocity dispersions may be lower than CDM allows for.

ACKNOWLEDGEMENTS

We would like to thank Mikko Laine for supplying the code that calculates the sterile neutrino distribution functions, Wojtek Hellwing for use of the COCO simulation data and Carlton Baugh for the use of the Planck-tuned GALFORM models. We would also like to thank John Helly for providing the subhalo accretion times in GALFORM. CSF acknowledges ERC Advanced Grant 267291 'COSMIWAY'. This work used the DiRAC Data Centric system at Durham University, operated by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). This equipment was funded by BIS National E-infrastructure capital grant ST/K00042X/1, STFC capital grant ST/H008519/1 and STFC DiRAC Operations grant ST/K003267/1 and Durham University. DiRAC is part of the National E-Infrastructure. This work is part of the D-ITP consortium, a programme of the Netherlands Organization for Scientific Research (NWO) that is funded by the Dutch Ministry of Education, Culture and Science (OCW). This work was supported in part by an STFC rolling grant to the ICC. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (GA No. 694896).

REFERENCES

Asaka T., Shaposhnikov M., 2005, Phys. Lett. B, 620, 17
Benson A. J., Frenk C. S., Lacey C. G., Baugh C. M., Cole S., 2002, MNRAS, 333, 177
Benson A. J. et al., 2013, MNRAS, 428, 1774
Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440
Bose S., Hellwing W. A., Frenk C. S., Jenkins A., Lovell M. R., Helly J. C., Li B., 2016, MNRAS, 455, 318
Boyarsky A., Ruchayskiy O., Shaposhnikov M., 2009, Annu. Rev. Nucl. Part. Sci., 59, 191
Boyarsky A., Ruchayskiy O., Iakubovskyi D., Franse J., 2014, Phys. Rev. Lett., 113, 251301
Boyarsky A., Franse J., Iakubovskyi D., Ruchayskiy O., 2015, Phys. Rev. Lett., 115, 161301
Boylan-Kolchin M., Bullock J. S., Kaplinghat M., 2011, MNRAS, 415, L40
Boylan-Kolchin M., Bullock J. S., Kaplinghat M., 2012, MNRAS, 422, 1203
Boylan-Kolchin M., Bullock J. S., Sohn S. T., Besla G., van der Marel R. P., 2013, ApJ, 768, 140
Bozek B., Boylan-Kolchin M., Horiuchi S., Garrison-Kimmel S., Abazajian K., Bullock J. S., 2016, MNRAS, 459, 1489
Brooks A. M., Zolotov A., 2014, ApJ, 786, 87
Bryan G. L., Norman M. L., 1998, ApJ, 495, 80
Bulbul E., Markevitch M., Foster A., Smith R. K., Loewenstein M., Randall S. W., 2014, ApJ, 789, 13
Bullock J. S., Kravtsov A. V., Weinberg D. H., 2000, ApJ, 539, 517
Busha M. T., Wechsler R. H., Behroozi P. S., Gerke B. F., Klypin A. A., Primack J. R., 2011, ApJ, 743, 117
Caldwell N. et al., 2016, ApJ, preprint (arXiv:1612.06398)
Campbell D. J. R. et al., 2015, MNRAS, 452, 852
Campbell D. J. R. et al., 2016, preprint (arXiv:1603.04443)
Cautun M., Frenk C. S., van de Weygaert R., Hellwing W. A., Jones B. J. T., 2014, MNRAS, 445, 2049
Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168
Colín P., Valenzuela O., Avila-Reese V., 2008, ApJ, 673, 203
Collins M. L. M. et al., 2013, ApJ, 768, 172
Correa C. A., Wyithe J. S. B., Schaye J., Duffy A. R., 2015a, MNRAS, 450, 1514
Correa C. A., Wyithe J. S. B., Schaye J., Duffy A. R., 2015b, MNRAS, 450, 1521
Correa C. A., Wyithe J. S. B., Schaye J., Duffy A. R., 2015c, MNRAS, 452, 1217
Cyr-Racine F.-Y., Sigurdson K., Zavala J., Bringmann T., Vogelsberger M., Pfrommer C., 2016, Phys. Rev. D, 93, 123527
de Vaucouleurs G., de Vaucouleurs A., Corwin H. G. Jr, Buta R. J., Paturel G., Fouqué P., 1991, Third Reference Catalogue of Bright Galaxies. Springer, New York, RC3
Deason A. J., Belokurov V., Evans N. W., An J., 2012, MNRAS, 424, L44
Di Cintio A., Knebe A., Libeskind N. I., Brook C., Yepes G., Gottloeber S., Hoffman Y., 2013, MNRAS, 431, 1220
Diemand J., Kuhlen M., Madau P., 2007, ApJ, 657, 262
Dodelson S., Widrow L. M., 1994, Phys. Rev. Lett., 72, 17
Dolgov A., Hansen S., 2002, Astropart. Phys., 16, 339
Fattahi A., Navarro J. F., Sawala T., Frenk C. S., Sales L. V., Oman K., Schaller M., Wang J., 2016, MNRAS, preprint (arXiv:1607.06479)
Garzilli A., Boyarsky A., Ruchayskiy O., 2015, preprint (arXiv:1510.07006)
Ghiglieri J., Laine M., 2015, J. High Energy Phys., 2015, 171
Gilmore G., Wilkinson M. I., Wyse R. F. G., Kleyna J. T., Koch A., Evans N. W., Grebel E. K., 2007, ApJ, 663, 948
González R. E., Kravtsov A. V., Gnedin N. Y., 2013, ApJ, 770, 96
Guo Q., Cooper A. P., Frenk C., Helly J., Hellwing W. A., 2015, MNRAS, 454, 550
Guo Q. et al., 2016, MNRAS, 461, 3457 (G16)
Hellwing W. A., Frenk C. S., Cautun M., Bose S., Helly J., Jenkins A., Sawala T., Cytowski M., 2016, MNRAS, 457, 3492
Irwin M., Hatzidimitriou D., 1995, MNRAS, 277, 1354
Jenkins A., Frenk C. S., White S. D. M., Colberg J. M., Cole S., Evrard A. E., Couchman H. M. P., Yoshida N., 2001, MNRAS, 321, 372
Jiang F., van den Bosch F. C., 2015, MNRAS, 453, 3575
Jiang F., van den Bosch F. C., 2016, MNRAS, 458, 2848
Kahn F. D., Woltjer L., 1959, ApJ, 130, 705
Kennedy R., Frenk C., Cole S., Benson A., 2014, MNRAS, 442, 2487
Kuhlen M., 2010, Adv. Astron., 2010, 162083
Lacey C. G. et al., 2016, MNRAS, 462, 3854 (LC16)

Laine M., Shaposhnikov M., 2008, J. Cosmol. Astropart. Phys., 6, 31
Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473
Li Y., White S. D. M., 2008, MNRAS, 384, 1459
Lovell M. R. et al., 2012, MNRAS, 420, 2318
Lovell M. R., Frenk C. S., Eke V. R., Jenkins A., Gao L., Theuns T., 2014, MNRAS, 439, 300
Lovell M. R. et al., 2016a, preprint (arXiv:1611.00010)
Lovell M. R. et al., 2016b, MNRAS, 461, 60
Ludlow A. D., Bose S., Angulo R. E., Wang L., Hellwing W. A., Navarro J. F., Cole S., Frenk C. S., 2016, MNRAS, 460, 1214
Macciò A. V., Paduroiu S., Anderhalden D., Schneider A., Moore B., 2012, MNRAS, 424, 1105
Macciò A. V., Paduroiu S., Anderhalden D., Schneider A., Moore B., 2013, MNRAS, 428, 3715
Martin N. F., de Jong J. T. A., Rix H.-W., 2008, ApJ, 684, 1075
McConnachie A. W., 2012, AJ, 144, 4
Navarro J. F., Eke V. R., Frenk C. S., 1996a, MNRAS, 283, L72
Navarro J. F., Frenk C. S., White S. D. M., 1996b, ApJ, 462, 563
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Navarro J. F. et al., 2010, MNRAS, 402, 21
Parkinson H., Cole S., Helly J., 2008, MNRAS, 383, 557
Parry O. H., Eke V. R., Frenk C. S., Okamoto T., 2012, MNRAS, 419, 3304
Peñarrubia J., Navarro J. F., McConnachie A. W., 2008, ApJ, 673, 226
Peñarrubia J., Benson A. J., Walker M. G., Gilmore G., McConnachie A. W., Mayer L., 2010, MNRAS, 406, 1290
Peñarrubia J., Ma Y.-Z., Walker M. G., McConnachie A., 2014, MNRAS, 443, 2204
Peñarrubia J., Gómez F. A., Besla G., Erkal D., Ma Y.-Z., 2016, MNRAS, 456, L54
Piffl T. et al., 2014, A&A, 562, A91
Planck Collaboration XVI, 2014, A&A, 571, A16
Polisensky E., Ricotti M., 2014, MNRAS, 437, 2922
Pontzen A., Governato F., 2012, MNRAS, 421, 3464
Ruchayskiy O. et al., 2016, MNRAS, 460, 1390
Sales L. V., Navarro J. F., Abadi M. G., Steinmetz M., 2007a, MNRAS, 379, 1464
Sales L. V., Navarro J. F., Abadi M. G., Steinmetz M., 2007b, MNRAS, 379, 1475
Sawala T., Frenk C. S., Crain R. A., Jenkins A., Schaye J., Theuns T., Zavala J., 2013, MNRAS, 431, 1366
Sawala T. et al., 2016a, MNRAS, 456, 85
Sawala T. et al., 2016b, MNRAS, 457, 1931
Schaller M. et al., 2015, MNRAS, 451, 1247
Schewtschenko J. A., Baugh C. M., Wilkinson R. J., Boehm C., Pascoli S., Sawala T., 2016, MNRAS, 461, 2282
Schneider A., 2016, J. Cosmol. Astropart. Phys., 4, 059
Schneider A., Smith R. E., Reed D., 2013, MNRAS, 433, 1573
Schneider A., Anderhalden D., Macciò A. V., Diemand J., 2014, MNRAS, 441, L6
Shao S., Gao L., Theuns T., Frenk C. S., 2013, MNRAS, 430, 2346
Shi X., Fuller G. M., 1999, Phys. Rev. Lett., 82, 2832
Simha V., Cole S., 2016, MNRAS, preprint (arXiv:1609.09520)
Springel V. et al., 2008, MNRAS, 391, 1685
Strigari L. E., Frenk C. S., White S. D. M., 2010, MNRAS, 408, 2364
Strigari L. E., Frenk C. S., White S. D. M., 2014, preprint (arXiv:1406.6079)
van der Marel R. P., Kallivayalil N., 2014, ApJ, 781, 121
Venumadhav T., Cyr-Racine F.-Y., Abazajian K. N., Hirata C. M., 2016, Phys. Rev. D, 94, 043515
Viel M., Lesgourgues J., Haehnelt M. G., Matarrese S., Riotto A., 2005, Phys. Rev. D, 71, 063534
Viel M., Becker G. D., Bolton J. S., Haehnelt M. G., 2013, Phys. Rev. D, 88, 043502
Vogelsberger M., Zavala J., Loeb A., 2012, MNRAS, 423, 3740
Vogelsberger M., Zavala J., Cyr-Racine F.-Y., Pfrommer C., Bringmann T., Sigurdson K., 2016, MNRAS, 460, 1399
Walker M. G., Peñarrubia J., 2011, ApJ, 742, 20
Walker M. G., Mateo M., Olszewski E. W., Peñarrubia J., Wyn Evans N., Gilmore G., 2009, ApJ, 704, 1274
Walker M. G., Mateo M., Olszewski E. W., Peñarrubia J., Wyn Evans N., Gilmore G., 2010, ApJ, 710, 886
Wang J., Frenk C. S., Navarro J. F., Gao L., Sawala T., 2012, MNRAS, 424, 2715
Wang W., Han J., Cooper A., Cole S., Frenk C., Cai Y., Lowing B., 2015, MNRAS, 453, 377
Wang M.-Y., Strigari L. E., Lovell M. R., Frenk C. S., Zentner A. R., 2016, MNRAS, 457, 4248
Wolf J., Martinez G. D., Bullock J. S., Kaplinghat M., Geha M., Muñoz R. R., Simon J. D., Avedo F. F., 2010, MNRAS, 406, 1220
Zavala J., Vogelsberger M., Walker M. G., 2013, MNRAS, 431, L20

APPENDIX A: CHANGE IN GALAXY FORMATION MODEL

The galaxy formation model can influence the satellite luminosity function in two ways: regulating the luminosity of the galaxy that can be formed in a halo with a given formation history, and destroying satellites through mergers with the host galaxy. In this appendix, we discuss these effects by analysing three versions of the GALFORM model: LC16, G16 and G16-2. We start with luminosity functions for our three assumed halo masses (0.5, 1.0, 1.4 $\times$ 10^{12} $M_\odot$). As

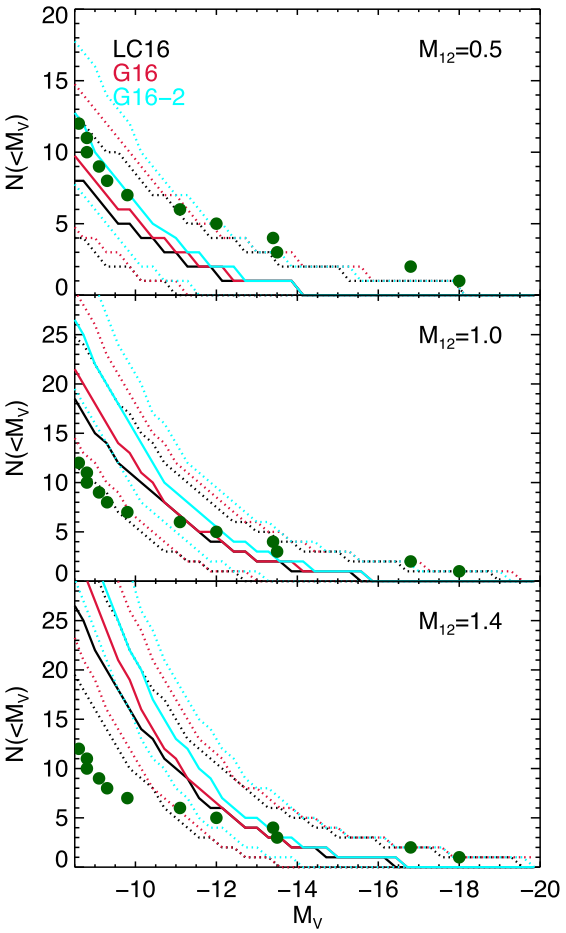

Figure A1. Cumulative satellite luminosity function for three ΛCDM galaxy formation models and three halo masses. The solid lines denote the median number of satellites brighter than M_V across all of our realizations, and the dotted curves mark the 9 and 95 percentiles. Each model is denoted by a different colour: LC16 (black), G16 (red) and G16-2 (cyan) as indicated in the legend. The top, middle and bottom panels show the mass functions for central haloes of mass 0.5, 1.0 and 1.4 $\times$ 10^{12} $M_\odot$ respectively. The circles mark the observed Milky Way satellite luminosity function.

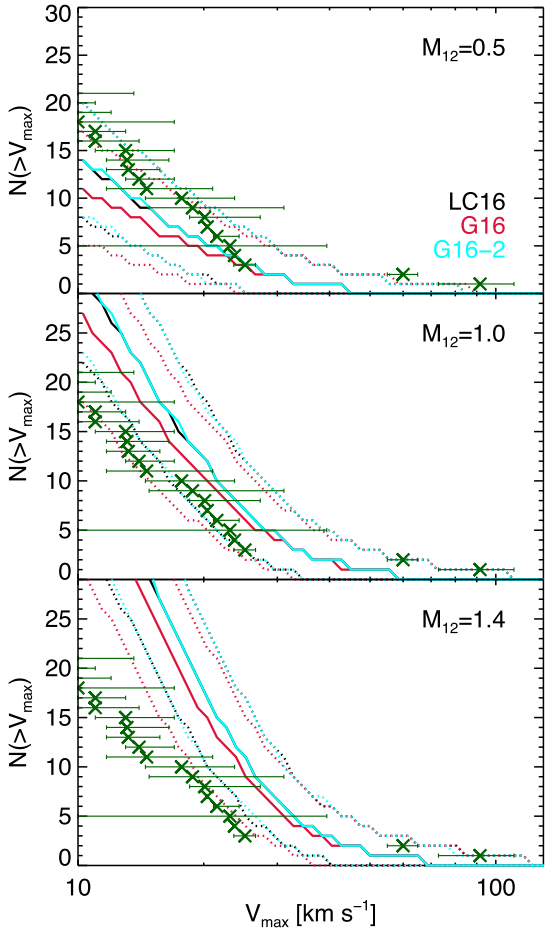

Figure A2. Cumulative satellite $V_{\rm max}$ function for the CDM versions of LC16 (black), G16 (red) and G16-2 (cyan). We include all luminous satellites, and have applied the baryonic physics feedback correction from equation (2). The solid lines denote the median $V_{\rm max}$ across all of our realizations, and the dotted curves again mark the 5 and 90 percentiles. The top, middle and bottom panels again show the mass functions for central haloes of mass 0.5, 1.0 and 1.4 $\times$ 10^{12} $M_\odot$ respectively. The dark green crosses mark the inferred Milky Way satellite $V_{\rm max}$ function assuming CDM.

seen in Fig. A1, the main difference amongst these models is that G16 produces more faint galaxies than LC16 due to its weaker feedback and G16-2 even more due to the lower merger rates of satellites on to the main galaxy (see below). As a result, the satellite luminosity functions are steeper than in LC16.

The $V_{\rm max}$ functions are plotted in Fig. A2. LC16 and G16-2 give nearly identical results, but for G16 the $V_{\rm max}$ function is slightly suppressed suggesting that haloes are more readily destroyed in this model. This explains the relative amplitudes of the luminosity functions in Fig. A1: LC16 suppresses the luminosity function through stronger feedback, and G16 through higher merger rates.

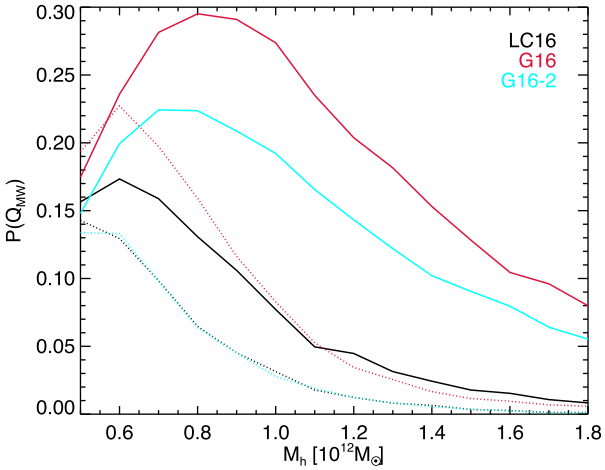

Figure A3. The probability that a Milky Way like satellite $V_{\rm max}$ distribution is drawn from the simulated distributions as a function of halo mass. The models used are LC16 (black), G16 (red) and G16-2 (cyan). Dotted lines denote results calculated from the base model (no luminosity-ordering or feedback suppression), and solid where luminosity-ordering and feedback suppression are included.

We now consider the implications for the probability of retrieving the Milky Way satellite luminosity function from the three model $V_{\rm max}$ distributions. We calculate $P(>Q_{\rm MW})$ as a function of halo mass and plot the results in Fig. A3. The base models (no luminosity-ordering or feedback suppression) of LC16 and G16-2 are almost identical, which results directly from their near-identical $V_{\rm max}$ functions. The elimination of some large satellites increases the amplitude of the G16 probability curve by up to a factor of 2 relative to G16-2. The boost to the likelihood introduced by including luminosity-ordering information is stronger for the G16 and G16-2 models, since their relatively low-mass satellites can form galaxies potentially as bright as their more massive counterparts. For G16 this model can generate Milky Way like satellite $V_{\rm max}$ functions in up to 30 per cent of realizations for low halo masses. We therefore conclude that greater scatter in the halo mass–luminosity relation may also play a role in removing the too big to fail problem. However, the models must be able to increase the scatter whilst simultaneously making all satellites fainter in order to fit the Milky Way satellite luminosity function.

This paper has been typeset from a TEX/LATEX file prepared by the author.